Molecular Cell, Volume *63*

Supplemental Information

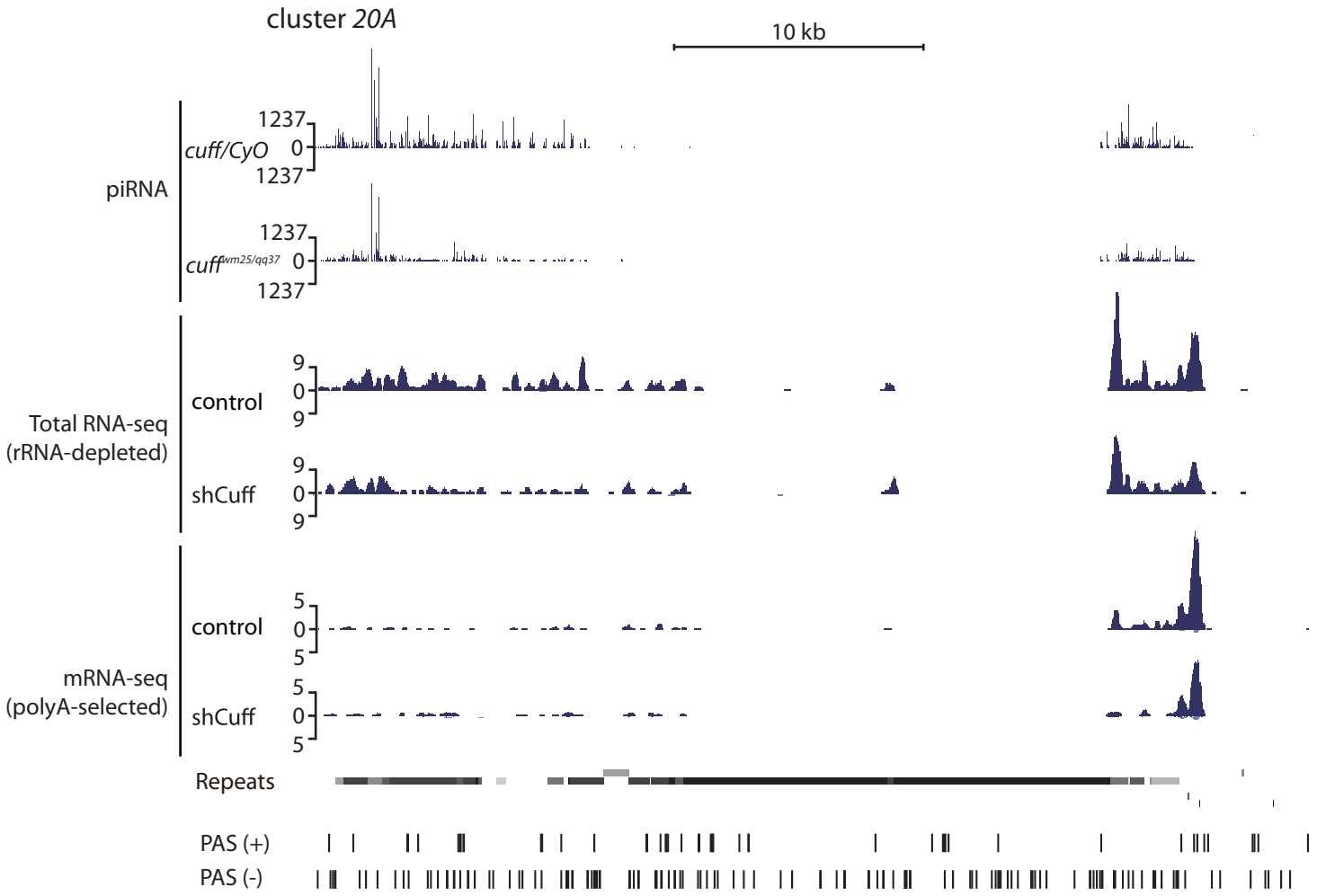
Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors

Yung-Chia Ariel Chen, Evelyn Stuwe, Yicheng Luo, Maria Ninova, Adrien Le Thomas, Ekaterina Rozhavskaya, Sisi Li, Sivani Vempati, John D. Laver, Dinshaw J. Patel, Craig A. Smibert, Howard D. Lipshitz, Katalin Fejes Toth, and Alexei A. Aravin

Inventory of Supplemental Material

Figure S1 Chen et al.

A



B

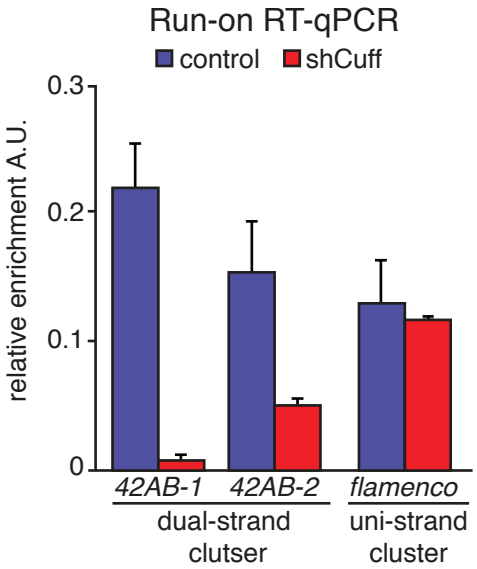
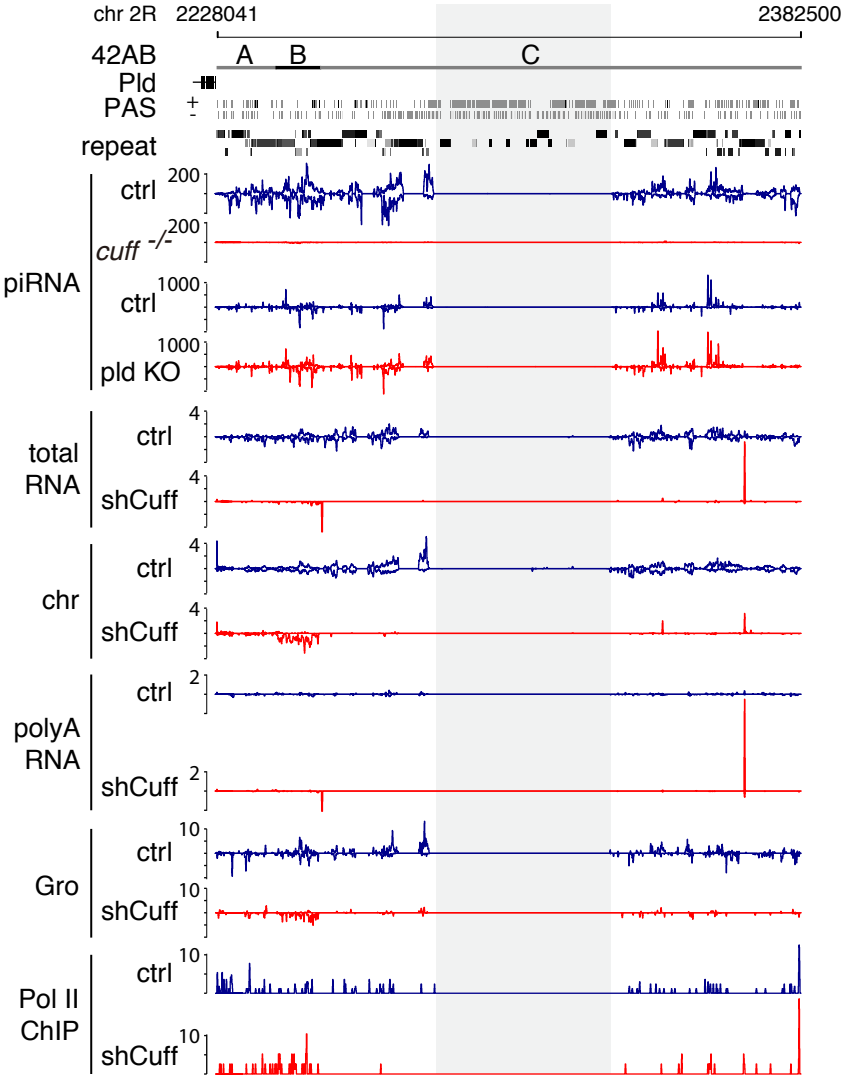
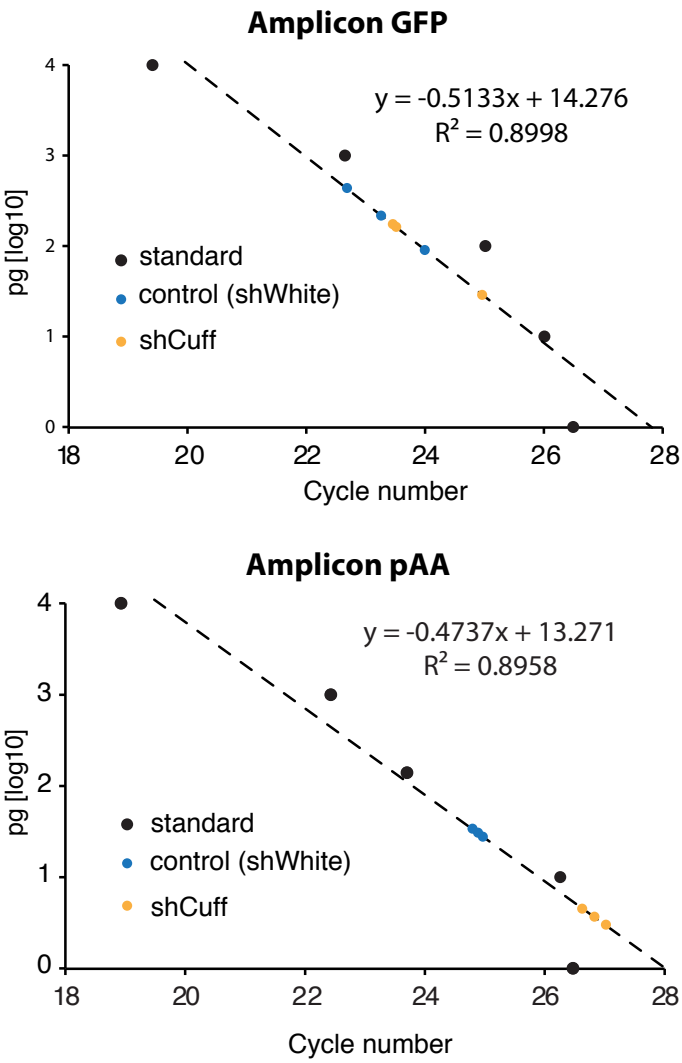


Figure S2 Chen et al.

A



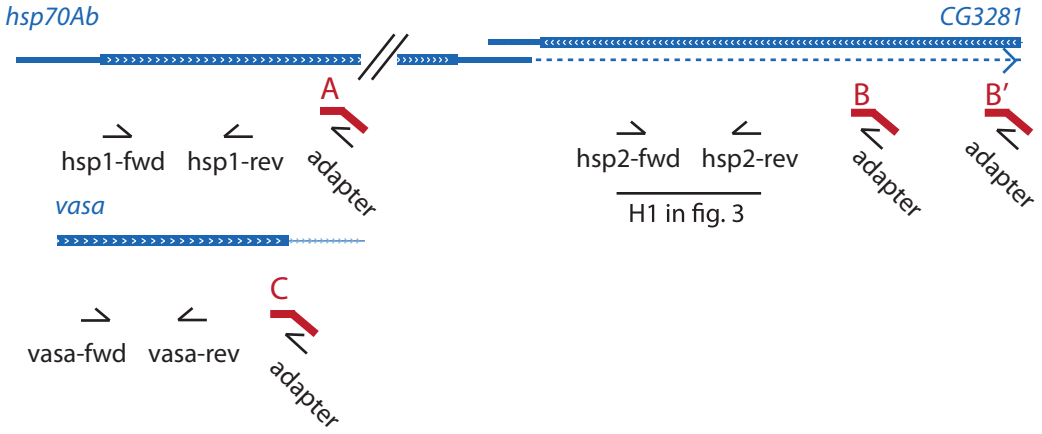
B



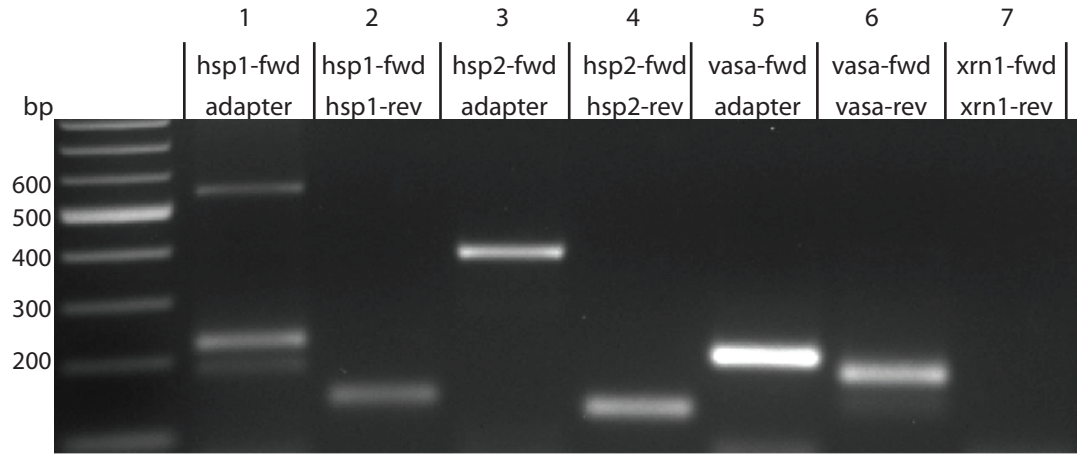
	GFP		pAA		
	Ct	mass [pg]	Ct	mass [pg]	pAA/GFP
control-1	24.0	91.1	25.0	28.3	31%
control-2	23.3	217.1	24.8	34.3	16%
control-3	22.7	428.3	23.7	113.8	27%
shCuff-1	25.0	29.1	26.8	3.6	13%
shCuff-2	23.5	160.6	27.0	2.9	2%
shCuff-3	23.5	171.4	26.6	4.6	3%

Figure S3 Chen et al.

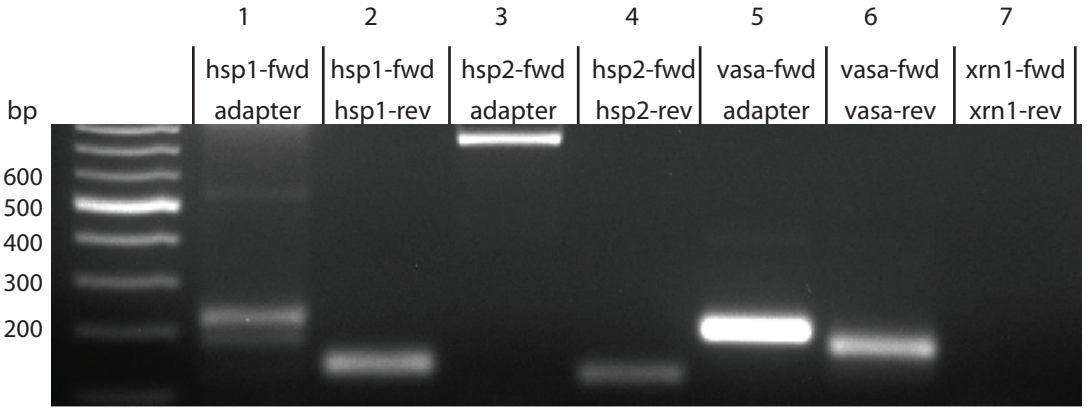
(A)



(B) RT with primers A,B,C



(C) RT with primers A,B′,C



(D) RT with random hexamers

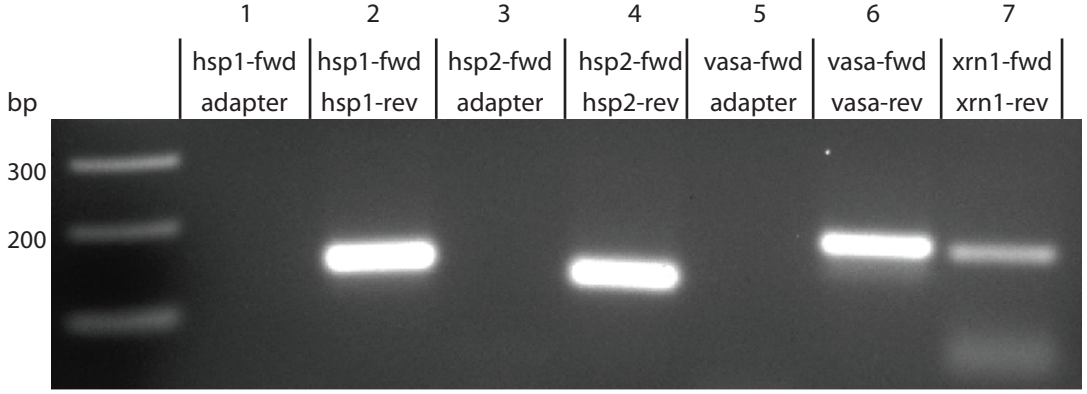
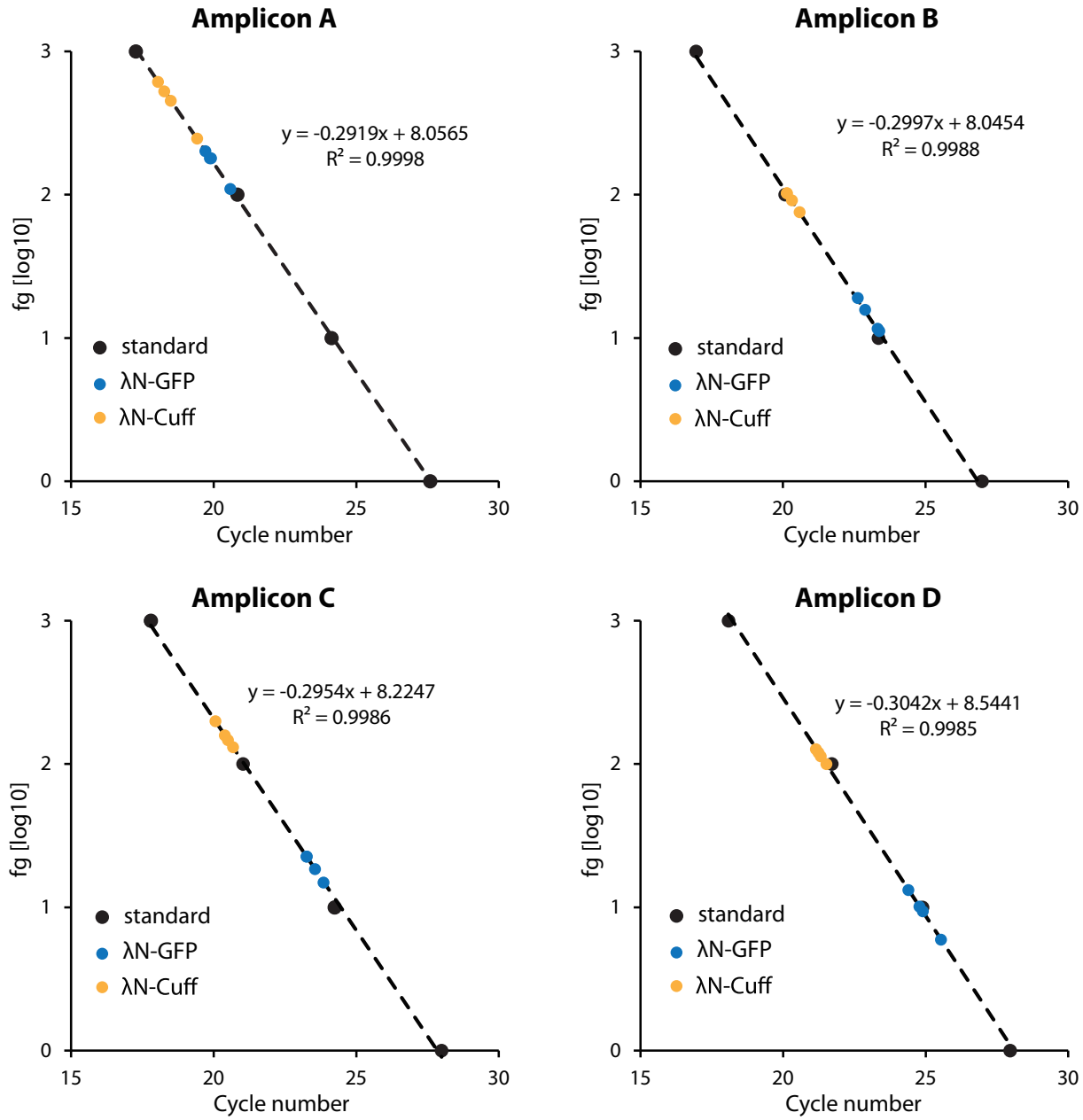


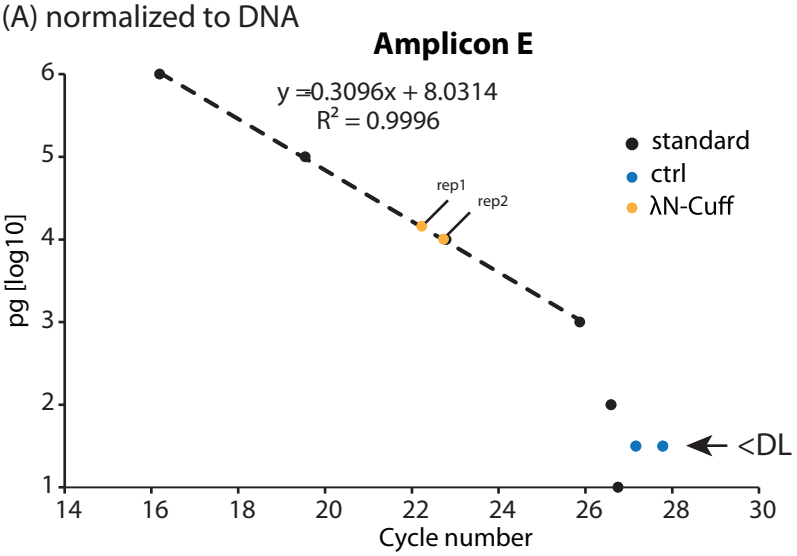
Figure S4 Chen et al.

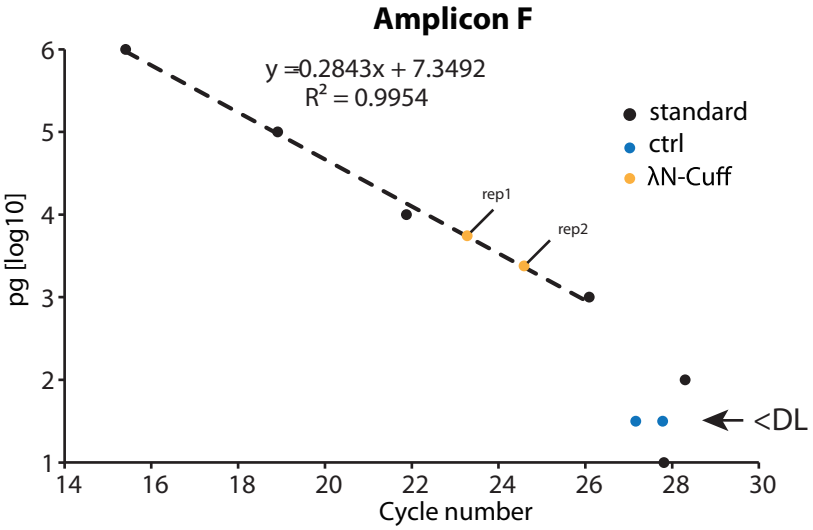


	A		B			C			D		
	Ct	mass (fg)	Ct	mass (fg)	B/A	Ct	mass (fg)	C/A	Ct	mass (fg)	D/A
λN-GFP-1	20.6	111.6	22.9	15.1	14%	23.5	18.6	17%	24.9	9.2	8%
λN-GFP-2	19.9	175.7	22.6	18.1	10%	23.3	22.7	13%	24.8	10.1	6%
λN-GFP-3	19.7	200.3	23.4	10.7	5%	23.8	15.2	8%	25.6	5.9	3%
λN-GFP-4	19.9	180.5	22.6	18.4	10%	23.2	22.9	13%	24.4	13.2	7%
λN-Cuff-1	18.1	611.3	20.1	102.6	17%	20.7	133.3	22%	21.5	100.5	16%
λN-Cuff-2	18.3	529.0	20.2	101.5	19%	20.5	150.6	28%	21.2	121.9	23%
λN-Cuff-3	19.4	243.4	20.6	75.2	31%	20.4	162.9	67%	21.3	115.6	48%
λN-Cuff-4	18.5	453.3	20.3	90.3	20%	20.0	203.9	45%	21.1	129.8	29%

Figure S5 Chen et al.

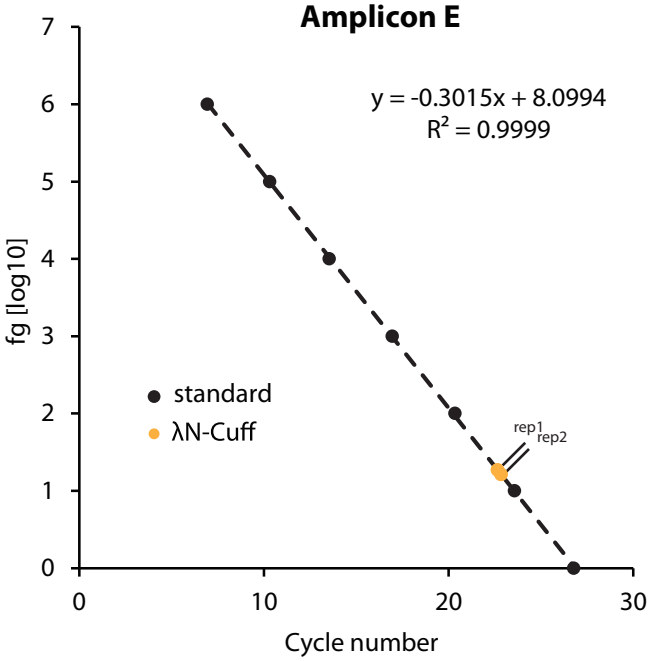
(A) normalized to DNA

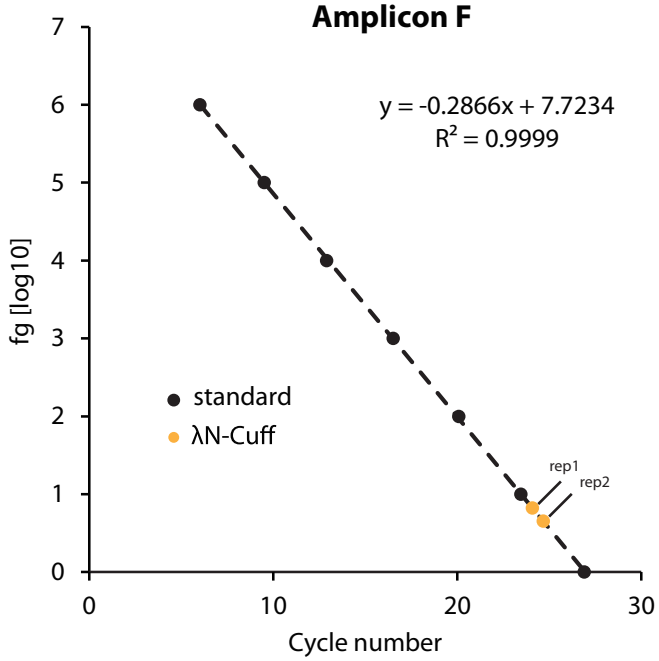




	E		F		
	Ct	mass (fg)	Ct	mass (fg)	F/E
λN-Cuff-1	22.2	14.1	23.3	5.4	38%
λN-Cuff-2	22.7	9.9	24.6	2.3	23%
ctrl-1	27.8	<DL	27.78	<DL	N/A
ctrl-2	27.2	<DL	27.16	<DL	N/A

(B) normalized to RNA





	E		F		
	Ct	mass (fg)	Ct	mass (fg)	F/E
λN-Cuff-1	22.7	18.6	24.1	6.7	36%
λN-Cuff-2	22.8	16.3	24.7	4.5	28%

Figure S6 Chen et al.

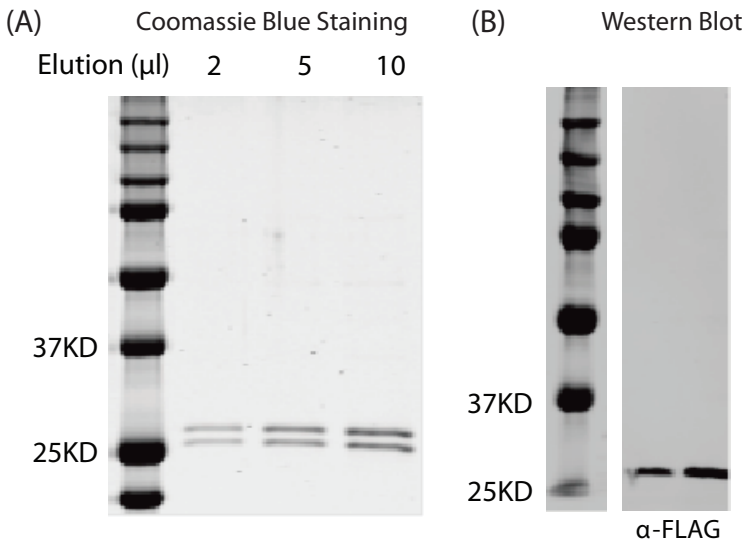
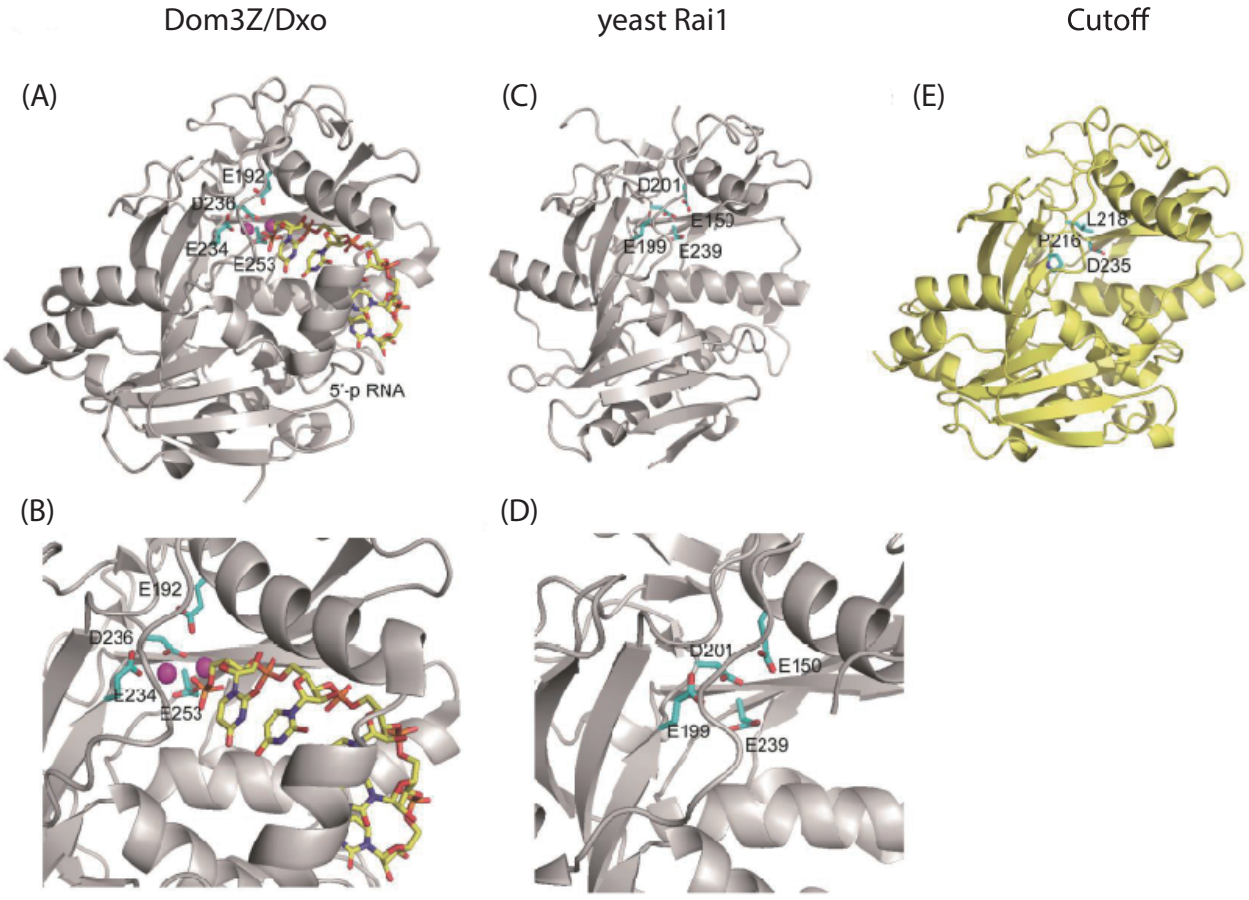


Figure S7 Chen et al.



Supplementary Figure Legends

Figure S1. The effect of Cuff on transcription of piRNA clusters, Related to Figure 1

(A) piRNA precursors from uni-strand cluster 20A are polyadenylated. Strand-specific profiles of piRNA, total and poly(A)-selected long RNA on uni-strand cluster *20A* in ovaries of control and Cuff-depleted (shCuff) flies. Y-axis shows RPM of uniquely mapped reads. Repeat densities and predicted canonical poly(A)-sites (PAS) on the negative and positive DNA strand are shown below.

(B) The effect of Cuff on transcription measured by nuclear run-on. Cuff depletion leads to decreased transcription on dual-strand clusters (such as *42AB*), but not on uni-strand clusters (like *flamenco*), as determined by nuclear run-on followed by RT-qPCR. Data was normalized to levels of *rp49*. Error bars represent the standard deviations of two biological replicates.

Figure S2. The effect of Cuff on *42AB* cluster, Related to Figure 2.

(A) Total and poly(A)-selected long RNA profiles on *42AB* cluster. Shown are uniquely mapped piRNA-seq, long RNA-seq, GRO-seq and RNA pol II ChIP-seq reads on the entire *42AB* clusters in ovaries of $cuff^{wm25/qq37}$ mutants and control heterozygous flies (cuff het) along with *cuff* (shCuff) and control (shWhite) shRNA-mediated knockdown flies. Only reads that are uniquely mapped to the genome are shown. RNA-seq signals were normalized to total library depth (RPM, reads per million) and density of the reads in 200-bp sliding windows is plotted.

(B) Quantification of MiMIC read-through transcripts. A plasmid containing the full sequence of the MiMIC transgene was serially diluted to make standard curves for amplification of the GFP and the pAA regions (Fig. 2A). For each amplicon, a trend line was generated, where the absolute DNA amount was plotted against the qPCR cycle number. This trend line was used to quantify the amount of cDNA produced by reverse transcription of RNA from control and cuff-depleted (shCuff) ovaries. For the RT, a mix of the strand-specific primers (shown in Fig. 2A and Table S1) was used. Three biological replicas for both samples are plotted on the graph. The cDNA quantification, based on the plasmid DNA amounts, is shown in the table. The ratio between pAA and GFP fragments in the plasmid used for the standard curves is 1:1. In contrast, the ratio is different for M8627 RNA, and these differences directly reflect the efficacy of termination upstream of the pAA region.

Figure S3. Strand-specific RT-PCR to detect *hsp70* read-through transcripts, Related to Figure 3.

(A) Schematic representation of the RT primers (red) containing a specific 5' portion and a universal adapter sequence in the 3' portion. PCR primers are shown in black.

(B) Reverse transcription was performed with primers A, B and C. Amplicons from the PCR reactions 2, 4, and 6 show that the transcripts of interest are detectable with specific primer pairs. Amplicons with the specific forward and the universal adapter reverse primer (lanes 1, 3, and 5) show that those primer pairs are suitable for detecting transcripts reverse transcribed from a specific strand. Reaction 3 detects an *hsp70* read through transcript. The PCR reaction with primer pair 7 on a control gene shows no product, demonstrating that reverse transcription is specific to RNA that matches the RT primer.

(C) Reverse transcription was performed with primers A, B' and C, the same conclusions as in panel B can be drawn. Since RT primer B' is downstream of B, the PCR amplicon is larger than in panel B.

(D) Reverse transcription was performed with random hexamers. The PCR reactions with specific forward and universal adapter reverse primers do not yield any products. This proves the specificity of the adapter reverse primer and shows that only transcripts from the desired strand are reverse transcribed. PCR with primer pair 7 shows a product, demonstrating that this primer is able to detect the control transcript if it is reverse transcribed with random hexamers.

Figure S4. Quantification of *mKate2-4BoxB* read-through transcripts, Related to Figure 4.

A plasmid containing the full sequence of the mKate2-4xBoxB reporter was serially diluted, and these samples were amplified with four primer pairs A to D (Fig. 4A). For each amplicon, a trend line was generated and used to enable quantification of the absolute amounts of cDNA produced by reverse transcription of RNA from ovaries of control and Cuff-tethered flies using random hexamer primer (four biological replicas). The results of cDNA quantification are shown in the table. The ratios among amplified regions when the plasmid was used as substrate were 1:1. In contrast, the ratios were different for reporter transcripts, indicative of the efficiency of transcription termination. For each of the B, C, and D amplicons the ratio relative to the level of amplicon A, which corresponds to mRNA region upstream of poly(A) site, was calculated.

Figure S5. Quantification of uncleaved *mKate2-4BoxB* read-through transcripts, Related to Figure 4.

(A) Quantification of uncleaved mKate2-4BoxB read-through transcripts. A plasmid containing the full sequence of the mKate2-4xBoxB reporter was serially diluted to make standard curves for PCR amplification of regions E and F (Fig. 4A). Region E is present in all read-through transcripts, whereas region F is present only in read-through transcripts that were not cleaved at the canonical poly(A) site. Therefore, the ratio between the levels of two fragments can be used to determine the fraction of unprocessed transcripts (F/E), and the fraction of cleaved transcripts is defined by subtracting unprocessed transcripts from all read-through transcript ((E-F)/E). For each amplicon, a trend line was generated and used to quantify the amount of cDNA produced by reverse transcription of ovarian RNA using a strand-specific primer (mKate-BoxB-K10polyA-D-R in Table S1) for the RT (two biological replicas). In control flies read-through transcripts levels fall out of the linear range of the standard curve and therefore are annotated as "below detection limit" (<DL). Results of the transcript quantification are shown in the table.

(B) To account for differences in the reverse transcription efficiency using specific RT primers, standard curves were generated with the same amplicons using RNA as template. An *in vitro* transcribed RNA, which contains regions E and F, was reverse-transcribed with the same specific primer as was used for RT from ovarian samples. The obtained cDNA was serially diluted to make the standard curves. The results of transcript quantification in Cuff tethering are shown in the table.

Figure S6. Purification of Fab antibody that binds CBP80 protein, Related to Figure 6.

(A) Coomassie blue staining of CBP80 antibody purified from *E. coli* cells.

(B) Western blot to detect FLAG-tagged antibody after purification.

Figure S7. Comparison of 5'-RNA recognition pockets, Related to Figure 7.

(A) Overall structure of 5'-p RNA bound DOM3Z (PDB code: 4J7L), with DOM3Z shown in silver ribbon and 5'-p RNA in stick representation. The residues involved in 5'-p recognition are highlighted by blue representation. **(B)** An enlarged view of the 5'-p RNA binding pocket in

DOM3Z to highlight the details of recognition. DOM3Z uses four acidic residues coordinated two magnesium ions (pink sphere) both for recognition of the 5'-p RNA and for catalytic activity. **(C)** Overall structure of yeast apo-Rai1 (PDB code: 3FQG) with Rai1 shown in silver ribbon and the residues corresponding to DOM3Z 5'-p RNA recognition highlighted by blue stick representation. **(D)** An enlarged view of the potential 5'-p RNA binding pocket in apo-Rai1 to show the details of recognition. **(E)** A model-based structure of Cutoff based on corresponding folds of DOM3Z and Rai1 is shown in a yellow ribbon representation. The residues corresponding to 5'-p RNA recognition and catalytic activity by DOM3Z that are not conserved between DOM3Z and Cutoff, are highlighted by blue stick representation.

Table S1. Oligonucleotide sequences, Related to Experimental Procedures

Dm_Rat1 shRNA-F	ctagcagtATCCGTGATAATAGAATGCAAtagttatattcaagcataTTGCATTCTATTATCACGGATgcg
Dm_Rat1 shRNA-R	aattcgcATCCGTGATAATAGAATGCAAtatgcttgaatataactaTTGCATTCTATTATCACGGATactg
Dm_Rai1 shRNA-F	ctagcagtTACGATAACTACAACTTAGAAtagttatattcaagcataTTCTAAGTTGTAGTTATCGTAgcg
Dm_Rai1 shRNA-R	aattcgcTACGATAACTACAACTTAGAAtatgcttgaatataactaTTCTAAGTTGTAGTTATCGTAactg
38C-F	TGACGGTCTCTATGGGCAGGC
38C-R	TTCAACAGCGACTGACTGCCG
GFP-RT	GTCCATGCCGAGAGTGATCC
GFP 1-F	GCTACCCCGACCACATGAAG
GFP 1-R	TCTTGTAGTTGCCGTCGTCC
GFP 2-F	CTGAAGGGCATCGACTTCAAGGA
GFP 2-R	ATAGACGTTGTGGCTGTTGTAGT
GFP 3-F	AACTACAACAGCCACAACGTCTA
GFP 3-R	GATCTTGAAGTTCGCCTTGATGC
GFP 4-F	CTATATCATGGCCGACAAGCAGA
GFP 4-R	CTTCTCGTTGGGGTCTTTGCTCA
pAA-RT	CGGGCTGCGTTCGAAATTTAT
pAA-F	GTTTCAGGTTCAGGGGGAGG
pAA-R	CTCAATGTGGTAATCGGGCG
RTprimer A (hsp70 specific)	tcaggattgatggtgcctacagcCTCCGCCGTCTCCTTCATCTTGGTCAGTA
RTprimer B (downstream hsp70)	tcaggattgatggtgcctacagcCCAAGCCCACAAACTGACAGCGG
RTprimer B' (downstream hsp70)	tcaggattgatggtgcctacagcGAAGTTGCTGCCAAGCCGGATG
RTprimer C (vasa specific)	tcaggattgatggtgcctacagcCTATTAGCTTGGATCGCTTGGCGTAC
hsp1-fwd	GATCTGGGCACCACCTACTC
hsp1-rev	GTTCTTAGCCGGATCGCCG
hsp2-fwd	GGTGTGCTGACGCATGTGAAG
hsp2-rev	CTTTCCCGTATCTAGCCTTACCATCC
vasa-fwd	CCAGAACATCAGACATTGATGTTTTCCGC
vasa-rev	CTATTAGCTTGGATCGCTTGGCGTAC
xrn1-fwd	GCGAGGGAATCACGCTAAACG
xrn1-rev	GAGTCGCGCACCATCAATGC
adapter	tcaggattgatggtgcctacagc
T1-F	ATCCGTCGACCTGCAGCCAAG
T1-R	CGTGGCATGCAATGCATTCTG
H1	CGTCTGTCCGCGTGTCTTTG
H2	CGTTTCTTCGGGTTCCAATGCG
mKate-BoxB-K10polyA-A-F	TCAGAGGGGTGAACTTCCCA
mKate-BoxB-K10polyA-A-R	CTCCCAGCCGAGTGTTTTCT
mKate-BoxB-K10polyA-B-F	AAACACTTTCCCCATCCCCG
mKate-BoxB-K10polyA-B-R	GCCCGAGTGTAATGATCCCC
mKate-BoxB-K10polyA-C-F	TACACTCGGGCCTACGGTTA
mKate-BoxB-K10polyA-C-R	CGATTACCAGTGATTGCGCT
mKate-BoxB-K10polyA-D-F	CGGCTTGATAGTGCAGGCTA
mKate-BoxB-K10polyA-D-R	CAGAGAAACGCAGGCGGATA
mKate-BoxB-K10polyA-E-F	GCAGCTCCATGAGATTCGCT
mKate-BoxB-K10polyA-E-R	ATACTGGCAAAGGTGCGGAA
mKate-BoxB-K10polyA-F-F	(same as mKate-BoxB-K10polyA-B-F)
mKate-BoxB-K10polyA-F-R	(same as mKate-BoxB-K10polyA-E-R)
mKate-BoxB-K10polyA-PCR1/2-F	GCACTCGAGTACTAAACTGGG
mKate-BoxB-K10polyA-PCR1-R	(same as mKate-BoxB-K10polyA-C-R)
mKate-BoxB-K10polyA-PCR2-R	(same as mKate-BoxB-K10polyA-E-R)
mKate-BoxB-K10polyA-mK2-5'-F	ATGGTGAGCGAGCTGATTAAGG
mKate-BoxB-K10polyA-mK2-5'-R	TGAAGTGGTGGTTGTTCACGG

mKate-BoxB-K10polyA-1-F	CCGTGAACAACCACCACTTC
mKate-BoxB-K10polyA-1-R	GTAGCCAGGATGTCGAAGGC
mKate-BoxB-K10polyA-2-F	GGGCTTCACATGGGAGAGAG
mKate-BoxB-K10polyA-2-R	CTCCCAGCCGAGTGTTTTCT
mKate-BoxB-K10polyA-3-F	GATGCCCGGCGTCTACTATG
mKate-BoxB-K10polyA-3-R	TGCTAGGGAGGTCGCAGTAT
mKate-BoxB-K10polyA-4-F	GCAATAGAACCCATGGATTTGA
mKate-BoxB-K10polyA-4-R	ACGTAAAGCAGGCTTGAAACG
mKate-BoxB-K10polyA-5-F	TTCCCAGCCACTTAAGCGAC
mKate-BoxB-K10polyA-5-R	TTGTCTCGGTTACCTTTGGT
Rp49-RT	CGGGAATATAATCGCAGCAGTTTT
Splicing reporter EGFP-RT	TGCTCAGGTAGTGGTTGTCG
Splicing reporter EGFP-right	GAACTTCAGGGTCAGCTTGC
Splicing reporter EGFP-spliced	ATATGGTGAGCAAGGGCGA
Splicing reporter EGFP-unspliced	CTCATCCACAGGTGAGCAAG
Flam-F	GATTACCATTTGGCTATGAGGATC
Flam-R	GCCCACAGACAAGCTACACA
rp49-F	CCGCTTCAAGGGACAGTATCTG
rp49-R	ATCTCGCCGCAGTAAACGC
42AB-5-F	GCTTTGGAGGAAGGGAGACAC
42AB-5-R	GCATCCAGCCAATTTACCCTC
42AB-6-F	CTTCAGCTTGGAGGAAACCAG
42AB-6-R	CAGCGACTTAAAGAGTGCAGTG
42AB-1-F	CGTCCCAGCCTACCTAGTCA
42AB-1-R	ACTTCCCGGTGAAGACTCCT
42AB-2-F	CTATTATTGGCACTGCTATCC
42AB-2-R	GGACCAATTAGCGCGAAGAC
42AB-3-F	CGCTGTTGAAAGCAAATTGA
42AB-3-R	GAGACCTTCGCTCCAGTGTC
42AB-4-F	GTGGAGTTTGGTGCAGAAGC
42AB-4-R	AGCCGTGCTTTATGCTTTAC
20A-A-F	GCCTACGCAGAGGCCTAAGT
20A-A-R	CAGATGTGGTCCAGTTGTGC
HetA-F	ATCCTTCACCGTCATCACCTTCCT
HetA-R	GGTGCGTTTAGGTGAGTGTGTGTT
Blood-F	TGCCACAGTACCTGATTTCG
Blood-R	GATTCGCCTTTTACGTTTGC
Gypsy-F	GTTCATACCCTTGGTAGTAGC
Gypsy-R	CAACTTACGCATATGTGAGT
Burdock-F	AGGGAAATATTTGGCCATCC
Burdock-R	TTTTGGCCCTGTAAACCTTG
TAHRE-F	CTGTTGCACAAAGCCAAGAA
TAHRE-R	GTTGGTAATGTTCGCGTCCT
TART-F	AGAGAGGGAAAGAAGGGAAAGGGA
TART-R	ATTTCCTGCCTGGTTAGATCGCCA
ZAM-F	ACTTGACCTGGATACACTCACAAC
ZAM-R	GAGTATTACGGCGACTAGGGATAC
gRNA1a	CTTCGTGTGTAACAGACAGGTAGT
gRNA1b	AAACACTACCTGTCTGTTACACAC
gRNA2a	CTTCGGCTACATGGCTTCGCGAGA
gRNA2b	AAACTCTCGCGAAGCCATGTAGCC
Pld-del-F	CTTAAGGTCAAGCCGAGCATACC
Pld-del-R	CGCGTGCTCTTTATATTCGTTGG

Table S2. Genomic location of piRNA clusters and regions within *42AB* cluster, Related to Experimental Procedures

Coordinates	*Name*
chr2R: 2144349 - 2382500	*42AB*
chr2R: 2142772 - 2166967	*42AB* region A
chr2R: 2166967 - 2185333	*42AB* region B
chr2R: 2185333 - 2389511	*42AB* region C
chr2L: 20148652 - 20223262	*38C*
chr2L: 20104577 - 20116331	*38C up*
chr3L: 23281510 - 23310943	*80EF*
chr3LHet: 1991313 - 2160576	
chr3LHet: 238123 - 332969	
chr3LHet: 1402377 - 1521600	
chr3R: 27894707 - 27898741	
chr4: 1280083 - 1325377	
chrU: 5766708 - 5772171	
chrU: 4015849 - 4029971	
chrX: 21505666 - 21684449	*Flamenco*
chrX: 21391192 - 21431907	*20A*
chrX: 21759393 - 21844063	

Supplemental Experimental Procedures

Fly stocks

The strain P1152, that carries insertion of the *P{lArB}* construct in telomeric sequences of the X chromosome (site 1A) is described in (Roche and Rio, 1998). The strain BC69 that has insertion of the *P{A92}* construct at a euchromatic location on chromosome 2L (site 35B10-35C1) is described in (Lemaitre et al., 1993). Both stocks were a generous gift from S. Ronsseray.

The stocks with MiMIC insertions into *42AB* cluster, M8627 and M7308, were obtained from the Bloomington stock center (BDSC #44802 and #43121, respectively). MiMIC transgene contains a promoterless GFP reporter followed by the SV40 3' UTR, which has a strong poly(A) signal (PAS).

cuff^{wm25} and *cuff*^{qq37} mutants are described in (Pane et al., 2011) and were generous gifts from T. Schupbach. The stocks for shRNA-mediated knock-down of *cuff* (shCuff, BDSC #35182) and *white* (shWhite, BDSC #33623) were obtained from the Bloomington stock center. To obtain stocks for shRNA-mediated knock-down of Rat1 (shRat1) and Rai1 (shRai1) the short hairpin RNA sequences (listed in Supplementary Table S1) were cloned into the pValium22 vector (Ni et al., 2011) and integrated into the attP2 landing site (BDSC #25710).

λN-Cuff (UASp-λN-GFP-Cuff) and BoxB reporter (UASp-mKate2-4xBoxB-K10) were generated by P-element integration. Control λN-GFP (UASp-λN-GFP-eGFP) was integrated into the attP9A landing site (BDSC #9736 and #9732). The expression of all constructs used in the tethering experiments was driven by maternal alpha-tubulin67C-Gal4 (MT-Gal4) (BDSC #7063) and nos-Gal4 (BDSC #4937). For all the other cases, UASp transgenes were driven by nos-Gal4 (BDSC #4937). To express Flag-tagged Cpsf73 UASp-FH-CPSF73 construct was integrated into the attP9A landing site (BDSC #9736).

The splicing reporter and the UASp-LacI-Rhino and UASp-LacI flies described in (Zhang et al., 2014) were generously provided by W. Theurkauf. To obtain LacI-tagged proteins, UASp-LacI-Gateway destination vector was generated and the Deadlock and Cuff sequences were integrated by recombination using LR clonase (Invitrogen) to obtain UASp-LacI-Deadlock and UASp-LacI-Cuff constructs. Both constructs were integrated into the attP3B landing site (BDSC #9723). All flies in the assays were put on yeast for 2 to 3 days before ovary dissections and were 3 to 14 days old at the time of dissection.

CRISPR/Cas9-mediated deletion of pld promoters

Using the CRISPR/Cas9 method described in (Kondo and Ueda, 2013) and http://shigen.nig.ac.jp/fly/nigfly/cas9/index.jsp, we deleted ~4.8 kb region (chr2R 2,128,572 – 2,133,370) that contains two alternative *pld* promoters. The region was defined based on peaks of Pol II Ser5 (ChIP-seq) and encompasses both *pld* transcription start sites. The Cas9 Target Finder tool (http://shigen.nig.ac.jp/fly/nigfly/cas9/index.jsp) was used to generate two gRNA pairs for targeting the intended promoter region (gRNA1a/gRNA1b and gRNA2a/gRNA2b), and cloned into plasmid pBFv-U62B. The resulting construct was injected into *Drosophila* embryos (strain 9736) by Bestgene. Transformants containing the *vermillion* marker were selected and subsequently crossed to the Cas9 containing fly line, CAS-0001 (y^2 cho^2 v^1; attP40{nos-Cas9}/CyO), until male founders were obtained. Mutants were identified through genomic PCR screening using primer pair pld-del-F and pld-del-R listed in Supplementary Table S1.

Cuff tethering to reporter

Cuff-tethering assay was performed in flies that expressed BoxB reporter (pUASp-mKate2-BoxB-K10) and λN-Cuff (pUASp-λN-GFP-Cuff) in ovaries under control of MT-Gal4 driver. The control flies had the same genotype but expressed λN-GFP (pUASp-λN-GFP-eGFP) or did not express any λN-fused protein as indicated. RNA extraction and RT-qPCR were carried out as described below. Pol II and CPSF73 ChIP were performed from ovaries of flies containing the pUASp-mKate2-BoxB-K10, the Nanos-Gal4 driver, either pUASp-λN-GFP-Cuff or pUASp-λN-GFP-eGFP as a control, and in the case of CPSF73 ChIP pUASp-FH-CPSF73.

RNA-seq and small RNA-seq libraries

Ovary total RNA (10-15 µg) was either depleted of ribosomal RNA with the Ribo-Zero™ rRNA Removal Kit (Epicentre) or was poly(A) selected with the Dynabeads® mRNA Purification Kit (Ambion). Strand-specific RNA-seq libraries were made using the NEBNext Ultra Directional RNA Library Prep Kit (NEB #E7420). Libraries were sequenced on the Illumina HiSeq 2500 platform (50-bp and 150-bp runs). Small RNA-seq data from ovaries of *cuff*$^{wm25/qq37}$ and heterozygous controls were previously published GEO # GSE59608 (Le Thomas et al., 2014).

Nuclear run-on and GRO-seq

Nuclear run-on experiments were performed on nuclei isolated from ovaries of shCuff flies. Expression of shRNA against *cuff* was driven by Nos-Gal4. Nos-Gal4 flies without shRNA were used as a control. The nuclear run-on procedure was carried out as previously described (Shpiz et al., 2011) with slight modifications. 5'-Bromouridine-5'-triphosphate (BrUTP; Sigma, B7166) labeled RNA was filtered through Illustra MicroSpin G25 columns (27-5325-01) twice to remove unincorporated BrUTP. The nuclear run-on RNA was captured using anti-BrdU antibody (Sigma, 032M 4753) by incubation for 1 hour followed by incubation with Protein G beads (Dynabeads, Invitrogen, 1003D) for 1 hour. The immunoprecipitation procedure was sequentially performed three times to yield highly enriched BrUTP-labeled RNA. As a negative control the same procedure was performed on non-labeled, total *Drosophila* ovary RNA. RT-qPCR was performed on 10% of the purified RNA with primers for *vasa*, *rp49*, and selected piRNA cluster regions. For global nuclear run-on (GRO-seq) (Core et al., 2008) immunoprecipitated RNA was used to clone total RNA libraries using the commercially available kit NEBNext Ultra Directional RNA Library Prep Kit (NEB #E7420). Libraries were sequenced on the Illumina HiSeq 2500 (SE 50 bp reads) platform.

Isolation of chromatin-bound RNA; chromatin-bound RNA library preparation

Isolation of nascent transcripts was performed as previously reported by Khodor and colleagues (Khodor et al., 2011), with some modifications. Flies were kept on yeast for 1-2 days. Ovaries were removed and were washed twice in PBS and then transferred into 500 µl buffer AT (15 mM HEPES-KOH pH 7,6, 10 mM KCl, 5 mM $Mg(CH_3COO)_2$, 3 mM $CaCl_2$, 300 mM sucrose, 0.1 % Triton X-100, 1 mM DTT, 1x protease inhibitor). After douncing 30 times with a tight pestle, the lysate was filtered through a miracloth sieve and the flow-through was pipetted onto 1 ml of Buffer B (15 mM HEPES-KOH pH 7,6, 10 mM KCl, 5 mM $Mg(CH_3COO)_2$, 3 mM $CaCl_2$, 1 M sucrose, 0.1 % Triton X-100, 1 mM DTT, 1x protease inhibitor). The sample was centrifuged at 5,900 x g for 15 min at 4 °C. The pellet was resuspended in 5 volumes nuclear lysis buffer (10 mM HEPES-KOH, pH 7.6, 100 mM KCl, 0.1 mM EDTA, 10 % glycerol, 0.1 mM $ZnCl_2$, 0.15 mM spermine, 0.5 mM sperimidine, 10 mM NaF, 0.2 mM Na_3VO_4, 1 mM DTT, 1x protease inhibitor, 1 U/µl RNasin plus) and dounced 10 times with a tight pestle. The lysate was put on ice and an equal volume 2X NUN (50 mM HEPES-KOH, pH 7.6, 600 mM NaCl, 2 M urea, 2% NP-40, 1x protease inhibitor) buffer was added drop by drop and vortexed. The sample was incubated on ice for 20 min and centrifuged at 16,100 x g for 30 min at 4 °C. After a wash in a 1:1 solution of nuclear lysis buffer and 2X NUN buffer the pellet was resuspended in Ribozol (Amresco, N580), and RNA was extracted following the manufacturer's instructions. Libraries

were prepared using the commercially available kit NEBNext Ultra Directional RNA Library Prep Kit (NEB #E7420). Sequencing was performed on the Illumina HiSeq 2500 (SE 100 bp reads) platform, and data was been previously published GEO # GSE79325 (Hur et al., 2016).

ChIP-seq and ChIP-qPCR

All the ChIP experiments followed the protocol described previously (Le Thomas et al., 2014). RNA Pol II ChIP was performed using anti-RNA Pol II from Abcam (ab5408). Rhi and Cuff ChIPs were carried out using a commercially available anti-GFP antibody (DHSB-4C9) (Sanchez et al., 2014). CBP80 ChIP was performed with CBP80 synthetic antibody prepared as described below. The CPSF ChIP was performed using anti-FLAG magnitude beads (Sigma). qPCR was performed on a Mastercycler® ep realplex PCR thermal cycler machine (Eppendorf). ChIP-qPCR data for CBP80 ChIP enrichment (Figure 6E) was calculated as IP to Input ratios of target region, normalized to IP to Input ratios for the *rp49* locus. PCR primers are listed in Table S1.

ChIP-seq library construction was done using the NEBNext ChIP-Seq Library Prep Master Mix Set (#E6240). Libraries were sequenced on the Illumina HiSeq 2500 (SE 75-bp reads) platform. The resulting sequencing reads were mapped against the genome dm3 using Bowtie 0.12.7 (Langmead et al., 2009) with the following settings: "-v 0 –m 1".

RT-PCR and RT-qPCR

RNA was isolated from ovaries with Ribozol (Amresco, N580) and was treated with DNase (Invitrogen, 18068-015). Reverse transcription was carried out using Superscript III (Invitrogen) with random hexamer unless otherwise specified. oligo dT was used for priming the RT reaction when detecting poly-adenylated RNA in the *38C* piRNA cluster. qPCR was performed as described above.

For GFP reporter mRNA from MiMIC insertions in Figure 2C, reporter mRNA in tethering experiments (Figure 4B), reporter mRNA in cap-IP experiments (Figure 6B), CBP80 RIP experiments (Figure 6D), and *rat1* KD experiments (Figure 7E-G), target expression was normalized against *rp49* mRNA.

For assessing the levels read-through and cleaved transcripts downstream of the poly(A) sites of reporter in tethering experiments (Figure 4C,E), expression of regions of interest was calculated relative to reporter mRNA using standard curves based on serial dilutions of the

reporter plasmid (see panel A for amplicon locations and Fig. S6 for detailed description of the procedure and for standard curves). P-values were calculated using Student's *t-test*.

Strand-specific reverse transcription for detecting *hsp70a* read-through transcripts was performed with Superscript III at 60 °C with primers containing a gene-specific sequence at the 5' end and a universal adapter sequence at the 3' end. The two reverse transcription reactions contained several sequence-specific primers: primer A (*hsp70*), primer C (*vasa*), and either primer B or primer B' (specific for the sequence downstream of *hsp70*). For controls, reverse transcription was performed with random hexamers at 50 °C after 5 min incubation at room temperature. qPCRs were performed with a Taq DNA polymerase (GenScript, E00007-1000) with the primer pairs 1-fwd/1-rev, 2-fwd/2-rev, and 3-fwd/3-rev to detect the strand-specific amplicons. To ensure strand specificity of the reverse transcription we performed control PCRs (Fig. S4). The levels of read through transcripts were normalized to the level of *hsp70* mRNA. PCR primers are listed in Fig. S1.

Capped-RNA immunoprecipitation (Cap-IP)

To prepare capped and control RNAs, the pCI-neo polylinker (Promega) was amplified by PCR with a forward primer corresponding to the T7 promoter and a reverse primer consisting of the T3 promoter sequences containing 16 cytosines at the 5' end. The PCR product (108 bp) was used as template to transcribe RNA with T7 RNA polymerase. The resulting RNA (91-nt) contained 16 guanosines at the 3' end to stabilize the 3' terminus (Xiang et al., 2009). To make radiolabeled 5'-capped-RNA, *in vitro* transcribed 5'-pppRNA was capped by vaccinia capping enzyme (NEB) in the presence of [α-^{32}P] GTP, with or without SAM. To radiolabel RNA with a single 5' phosphate for the Rat1 exonuclease assay, the *in vitro* transcribed 5'-pppRNA was dephosphorylated by calf intestinal phosphatase (NEB) for 1 h at 37 °C and subsequently phosphorylated with T4 polynucleotide kinase (NEB) in the presence of [γ-^{32}P] ATP at 37 °C for 1 h. The RNA used as in cap-IP quality control is a 20-nt RNA oligo (purchased from IDT) radiolabeled with T4 polynucleotide kinase (NEB) and [γ-^{32}P] ATP. All RNAs were gel purified by denaturing PAGE.

Cap-IP was carried out as previously described with some modifications (Chang et al., 2012). Ovary RNA (5 µg) was incubated with 15 µl agarose-conjugated anti-2,2,7-trimethylguanosine antibody (Calbiochem, NA02A) in 100 µl binding buffer at room temperature for 1 h. The beads were then washed three times with 200 µl wash buffer. Methyl-capped mRNAs were eluted from the beads with 200 µl of 0.5 mg/ml of proteinase K in wash buffer for 30 min at 37 °C. The

isolated methyl-capped mRNAs were extracted with Ribozol (800 µl) for RT-qPCR. For cap-IP quality control, radiolabeled 5'-monophosphate RNA and 5'-capped RNA (with or without methyl group) were spiked into 5 µg of ovary RNA prior to IP. The extracted RNA was then resolved in a 6% PAGE gel. The gel was exposed to a PhosphorImager screen and scanned with a Storm 860 Molecular Imager.

Expression and purification of CBP80 antibody

The synthetic antibody against *Drosophila* CBP80 was developed using phage display as described (Laver et al., 2012; Laver et al., 2013; Laver et al., 2015). To express the antibody, starter culture was inoculated with single *E. coli* colony that contained plasmid encoding the antibody. The next day, 2xYT medium was inoculated by adding 1/100 volume of starter culture and grown at 37 °C to OD_{600} 0.6~0.8. Cells were cooled to 18 °C, and IPTG was added to final concentration 0.5 mM, followed by incubation of culture at 18 °C to allow antibody expression. After 24 h, cells were harvested by centrifuging at 4,000 rpm for 25 min, and cell pellet was lysed in Lysis Buffer (B-PER Bacterial Protein Extraction Reagent (Thermo Scientific), 1 µl/mL Lysis Buffer Benzonase nuclease (Novagen), 1 mM AEBSF, 2 mM benzamidine, 2 µg/mL leupeptin, 2 µg/mL pepstatin). The lysate was clarified by centrifugation at 15,000 rpm for 30 min at 4 °C. The supernatant was loaded on a column packed with Protein A beads (GE Life Sciences). The column was washed with Washing Buffer (20 mM Tris-HCl, pH 7.5, 500 mM NaCl), followed by wash with PBS. Antibody was eluted by addition of Elution Buffer (50 mM NaH_2PO_4, 100 mM H_3PO_4, 140 mM NaCl, pH 2.8), followed by addition of 3/10 volume of Neutralization Buffer (1 M Na_2HPO_4, 140 mM NaCl, pH 8.6). Eluted antibody were dialyzed with PBS overnight and stored at 4 °C.

CBP80 RIP

Fifty ovaries from flies expressing the mK2-4BoxB reporter and either λN-Cuff or the λN-GFP control were homogenized in lysis buffer (0.2% TritonX-100, 0.2% NP40, 150 mM NaCl, 5% glycerol, 50 mM Tris, pH 7.4, 2 mM DTT, proteinase inhibitor, RNase inhibitor) and clarified by centrifugation at 12,000 rpm. The supernatant was incubated with anti-FLAG magnetic beads (Sigma) that were conjugated with FLAG-tagged CBP80 antibody at 4 °C overnight. The beads were washed three times with Buffer A (0.1% NP40, 0.1% Triton X-100, 250 mM NaCl, 50 mM Tris-HCl, pH 7.4, 5% glycerol, proteinase inhibitor, RNase inhibitor) for 3 times, and three times

with Buffer B (PBS, 0.1% Triton X-100) for 3 times. RNA was extracted after Proteinase K digestion of the beads using TRIzol.

Drosophila Rat1 protein expression and purification

The *Drosophila Rat1* gene (1-768) was inserted into the pET28a expression vector which fuses a hexa-His tag at the C-terminus to the protein. The plasmid was transformed into *E. coli* BL21 (DE3) strain (Stratagene). The cells were cultured at 37 °C until OD_{600} reached 0.8, and then the protein expression was induced with 0.2 mM IPTG, and cells were incubated at 18 °C overnight. The hexa-His tagged protein was purified using a HisTrap FF column (GE Healthcare) and then further purified by a HiTrap Q FF column (GE Healthcare) and a Hiload Superdex G200 16/60 column (GE Healthcare).

Exoribonuclease assay

Exoribonuclease assays were performed at 25 °C for 1 hour. The reaction mixtures contained ~1 ng 5'-p*-RNA, 50 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris, pH 8.0, 0.5 mM DTT, 500 ng BSA, and the 500 nM Rat1. A reaction without Rat1 protein was as a control. The RNA products were isolated and fractionated by 6% PAGE in 7.5 M urea. The gel was exposed to a PhosphorImager screen and scanned with a Storm 860 Molecular Imager.

Pyrophosphohydrolase activity assay

Pyrophosphohydrolase assays was performed as described (Xiang et al., 2009). The indicated concentration of proteins (Figure 7B) were incubated with 1nM 5'-p*pp-RNA in RNA decapping buffer (10 mM Tris-HCl (pH 7.5), 100 mM KOAc, 2 mM MgOAc, 0.5 mM MnCl2 and 2 mM DTT) at 37°C for 30 min. The product was resolved on polyethyleneimine-cellulose thin layer chromatography plates developed in 1.5 M KH2PO4 (pH 7.5) followed by exposure to a PhosphorImager screen then scanned with a Storm 860 Molecular Imager. p*pi: pyrophosphate released from p*ppRNA.

Immunoprecipitation and western blot

For CBP80 immunoprecipitation (Figure 6C), the pActin-Myc-CBP80 construct was transfected into S2 cell using TransIT-LT1 (Mirus) 3 days prior to harvesting. IP was performed using anti-FLAG magnitude beads (Sigma) conjugated with FLAG-tagged anti-CBP80 antibody. Western blot was performed using anti-Myc-HRP antibody.

For co-IP experiments (Figure 7C) pcDNA6.2-dRat1-mKate2 with either pcDNA6.2-N-EmGFP-dRai1 or pDEST26-myc-GFP (control) plasmids were co-transfected into 293T cells with lipofectamine2000 (Life Technology). At 48 h after transfection, cells were lysed and immunoprecipitation was carried out using protein G magnetic Dynabeads (Invitrogen) conjugated with anti-mKate2 antibodies (AB233, Evrogen) or anti-GFP antibodies (rabbit polyclonal serum; Covance, affinity-purified in our laboratory). Western blotting was performed using the same antibodies.

Bioinformatic analysis

For all RNA-seq experiments, reads were first processed with Reaper tool, v. 12-205 (Davis et al., 2013) or cutadapt (Martin, 2011), to eliminate potential adaptor ("GATCGGAAGAGCACA" or "TGGAATTCTCGGGTC") contamination, and shorter than 15 nt reads were discarded. Next, reads were aligned to rRNA sequences (including rRNA unit, (Stage and Eickbush, 2007); 5S rRNA in dm3 annotated by RepeatMasker (http://www.reapeatmasker.org), as available via the UCSC Genome Bioinformatics Site, (Karolchik et al., 2014), using Bowtie 0.12.7 (Langmead et al., 2009) allowing up to three mismatches. Reads aligning to rRNA were also discarded from further analyses.

The remaining reads were aligned to dm3 with Bowtie 0.12.7 (Langmead et al., 2009), retaining only uniquely mapping reads. Bowtie alignment mismatch parameter (-v) was chosen differently according to read lengths, with 0, 2 or 3 mismatches allowed for 50, 100 and 150 bp libraries, respectively. For small RNA libraries, we allowed 1 mismatch. We note that changing the -v parameter within the 0-3 range does not affect our results (data not shown).

Read counts corresponding to regions of interest and downstream analyses were calculated using bedtools (Quinlan and Hall, 2010), SAM tools (Li et al., 2009), and custom perl, python and R scripts available upon request. Data was normalized as RPM (reads per million) or RPKM (reads per kilobase per million).

Genomic coverage histograms as in Figures 1A and 2A were generated for 200 bp sliding windows using bedtools (Quinlan and Hall, 2010), using only uniquely mapping reads and normalizing as reads per million (RPM). All RNA-seq libraries are directional with the orientation of the reads indicated. PolyA sites were annotated by scanning the dm3 genome for the AATAAA or complementary string using a custom script. The distribution of predicted poly(A) signals (AATAAA) on plus and minus genomic strands is shown in the poly(A) signal (PAS)

tracks, and distribution of transposable element sequences (UCSC repeat masker track) is shown in the repeats track. For piRNA analysis, *cuff*$^{wm25/qq37}$ mutant was compared to *cuff*/CyO heterozygotes; for all other experiments shRNA-mediated knockdown of *cuff* (shCuff) was compared to shRNA-mediated knockdown of *white* or strains containing identical driver but no hairpin (control). In Figure 1A, an expansion of the indicated region in the *38C* cluster from which poly-adenylated transcripts are produced in *cuff*$^{wm25/qq37}$ flies is shown. RNA was detected by RT-PCR using oligo-dT primer for reverse transcription and the PCR primers indicated on the figure. The RT-PCR product was cloned and sequenced to confirm its origin from the *38C* cluster.

For calculating read coverage of intronic and exonic regions, RefSeq gene table downloaded from the University of California at Santa Cruz [UCSC] genome browser (Meyer et al., 2013) was manually curated. Exons were defined as the region of the genome covered by any of the exons annotated by RefSeq (Pruitt et al., 2009), and introns were defined as regions covered by intron RefSeq annotations, but not exons. In the cases of tRNA, snoRNA, rRNA or other well known structured non-coding RNA annotations overlapping protein-coding gene annotations, reads corresponding to the regions of overlap were excluded before protein-coding gene expression estimation. In the case of overlapping protein-coding genes in the same orientation, reads corresponding to overlapping regions were evenly split between the annotated genes. Coordinates of the piRNA clusters are given in Supplementary Table S2.

To calculate intronic read enrichment in chromatin RNA-seq (Figure 1C), intronic to exonic read ratios were compared between chromatin RNA-seq and total RNA-seq libraries. We selected genes with at least one intronic read in the control ovaries, and intronic space larger than 10Kb.

To access piRNA cluster enrichment in chromatin RNA-seq data (Figure 1D), clusters were divided in 5kb windows. Ratios of RNA coverage in chromatin RNA-seq to total RNA-seq for 5kb windows spanning piRNA clusters was compared to these of exons and introns of different size ranges (0.1-1Kb, 1-10Kb, and larger than 10Kb). As many introns have no reads in standard RNA-seq data, value of 1 read was added to each data point to avoid division by 0.

For genome-wide comparisons of different RNA sequencing data in Cuff-depleted and control ovaries, dm3 genome was split in 5Kb (Figure 1E) or 1Kb (Figure 1F) intervals. Genomic windows that produce piRNAs in a Cuff-dependent manner shown in Figure 1E,F were defined as windows that produce piRNAs in control ovaries (10 RPKM cutoff), and which display greater than 80% piRNA reduction upon Cuff depletion. For each Cuff-dependent window, signal

change in Cuff-depleted versus control ovaries for different RNA-sequencing libraries (small RNA, total RNA, chromatin-associated RNA and GRO-seq) was log2-normalized. Heatmap of resulting values was generated using the 'gplots' R package. Windows are ordered by chromosome position, and windows overlapping previously annotated top piRNA clusters are marked. Rightmost column shows the piRNA strand bias for each 5 kB interval, defined as the proportion of piRNA originating from the less expressed DNA strand. Intervals overlapping uni-strand piRNA clusters (*flamenco* and *20A*, >=10 RPKM small RNA cutoff) that are not affected by Cuff depletion are shown for comparison (bottom). For chromatin, total and GRO-seq data, windows with less than 0.5 RPKM signal were excluded. For visualization purposes, counts of 1 were added to data in genomic windows before calculating log2 fold changes.

Cutoff modeling

Cutoff model structure was obtained from Phyre server based on the mouse Dom3Z protein (PDB code: 3FQI).

Accession codes

High-throughput sequencing data for RNA-seq, GRO-seq and ChIP-seq experiments are available through Gene Expression Omnibus (accession no. GSE81090), and small RNA sequencing data and chromatin RNA sequencing data were previously published and available through Gene Expression Omnibus (accession no. GSE59608 and GSE79325)

Supplemental References

Chang, J.H., Jiao, X., Chiba, K., Oh, C., Martin, C.E., Kiledjian, M., and Tong, L. (2012). Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. Nat Struct Mol Biol *19*, 1011-1017.

Davis, M.P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., and Enright, A.J. (2013). Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods *63*, 41-49.

Hur, J.K., Luo, Y., Moon, S., Ninova, M., Marinov, G.K., Chung, Y.D., and Aravin, A.A. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev *30*, 840-855.

Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M.*, et al.* (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res *42*, D764-770.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev *25*, 2502-2512.

Kondo, S., and Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. Genetics *195*, 715-721.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Laver, J.D., Ancevicius, K., Sollazzo, P., Westwood, J.T., Sidhu, S.S., Lipshitz, H.D., and Smibert, C.A. (2012). Synthetic antibodies as tools to probe RNA-binding protein function. Mol Biosyst *8*, 1650-1657.

Laver, J.D., Li, X., Ancevicius, K., Westwood, J.T., Smibert, C.A., Morris, Q.D., and Lipshitz, H.D. (2013). Genome-wide analysis of Staufen-associated mRNAs identifies secondary structures that confer target specificity. Nucleic Acids Res *41*, 9438-9460.

Laver, J.D., Li, X., Ray, D., Cook, K.B., Hahn, N.A., Nabeel-Shah, S., Kekis, M., Luo, H., Marsolais, A.J., Fung, K.Y.*, et al.* (2015). Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. Genome Biol *16*, 94.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C., Luo, Y., Sachidanandam, R., Toth, K.F.*, et al.* (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev *28*, 1667-1680.

Lemaitre, B., Ronsseray, S., and Coen, D. (1993). Maternal repression of the P element promoter in the germline of Drosophila melanogaster: a model for the P cytotype. Genetics *135*, 149-160.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B.*, et al.* (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res *41*, D64-69.

Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res *37*, D32-36.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Roche, S.E., and Rio, D.C. (1998). Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the Drosophila Polycomb group gene, Enhancer of zeste. Genetics *149*, 1839-1855.

Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M., and Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. Nucleic Acids Res.

Stage, D.E., and Eickbush, T.H. (2007). Sequence variation within the rRNA gene loci of 12 Drosophila species. Genome Res *17*, 1888-1897.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., and Tong, L. (2009). Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature *458*, 784-788.